

SYSTEM AND METHOD FOR AUTOMATICALLY REDUCING NOISE FOR VIDEO ENCODING

CROSS-REFERENCE TO RELATED APPLICATIONS

5 This application claims the benefit of U.S. Provisional Application, Serial No. 60/269,155, filed on February 15, 2001, and entitled "Automatic Noise Reduction for MPEG Video Encoding," which is incorporated by reference herein in its entirety.

FIELD OF THE INVENTION

10 The present invention generally relates to video encoding and, more particularly, is related to the reduction of noise for video encoding.

BACKGROUND OF THE INVENTION

With advancements in technology, significant advances have been made in video processing technology. Analog video, which provides limited compression through typical single scan-line or one-dimensional ("1-D") processing, has been surpassed by more efficient multiple scan-line or two-dimensional ("2-D") digital video processing. Two-dimensional digital video processing has been surpassed by horizontal, vertical and temporal or three-dimensional ("3-D") digital video processing. Even MPEG-1, which was once the predominant mainstream 3-D video codec standard, has also recently been surpassed by the more versatile and higher-bit-rate-capable MPEG-2. Presently, MPEG-2 is the predominant mainstream compression standard.

As is known in the art, video recording is susceptible to different degrees and categories of noise that negatively affects video during compression and encoding.

25 Samples of such noise include impulsive noise, such as, but not limited to, spikes, high

contrast glitches, and “salt-n-pepper effect.” Another sample of such noise includes Gaussian-distributed white noise, such as, but not limited to, thermal noise and “snow,” which may arise from cable ingress, interference in an analog section of a digitizer utilized during video encoding, or other phenomena.

5 Not only does noise degrade picture quality directly by virtue of its visual appearance, but the presence of noise degrades the quality of compressed video indirectly because, as is known in the art, an appreciable fraction of bit rate is consumed to compress the noise. This leads to two situations. First, less bit rate is readily available to compress the real signal, resulting in increased compression impairments. Second, as is known in
10 the art, compressed noise often has a more disturbing appearance than uncompressed noise. Thus, it is of primary importance to remove video noise prior to compression, particularly in situations where it is desired to minimize the bit rate.

 Digital video noise can be removed by a variety of known digital processing techniques such as, but not limited to, finite impulse response (FIR) linear spatial filters,
15 nonlinear spatial filters of various types, temporal filters, and even spatio-temporal filters. The temporal filter can take many forms, however, a typical temporal filter utilizes a motion-detector to moderate a recursive (infinite impulse response (IIR)) filter, which blends each pixel from a current image with a spatially co-located pixel from a previously filtered image. A large variety of noise reduction systems can be designed with different
20 configurations of these basic components. Such noise reduction systems typically are set to operate in a static mode, in the sense that the noise characteristics are assumed not to change over time. The settings are set based on user input or based on offline (not real-time) measurements or calculations of the digital video noise. However, in a real

environment, noise characteristics change over time or from video scene to video scene. Therefore, static noise reduction methods are insufficient, especially in systems such as consumer digital video recorders in which a user is not always present to control the recorder settings to compensate for changes in noise sources that may comprise widely
5 varying noise characteristics.

Furthermore, without additional controls, the classical motion-detecting temporal filter with a simple motion detector has difficulty separating moving objects from noise, and thus cannot always use the best noise reduction setting. Specifically, in certain scenes, the classical motion-detecting temporal filter will filter too little, leaving more noise than
10 necessary. Alternatively, in other scenes, the temporal filter may filter too much, visibly smearing moving objects or creating “ghosts,” which appear as attenuated copies of moving objects trailing behind objects as they move within a video.

SUMMARY OF THE INVENTION

In light of the foregoing, the preferred embodiment of the present invention generally relates to a system for providing automatic noise reduction for video encoding.

5 Generally, with reference to the structure of the noise reduction system, the system utilizes a video input module and a motion estimation unit. The video input module is capable of performing the steps of: filtering noise from currently received video data; combining the filtered data, wherein the combining step is dependent upon a category of the noise; and providing a weighted average of a current field derived from the combined
10 filtered data and a prior field, wherein the prior field is derived from previously combined and filtered data that has been previously stored, the weighted average being determined by pixel motion between the current field and the prior field.

 The motion estimation unit is capable of performing the steps of: separating a current video frame into multiple current regions of pixels and separating a prior video
15 frame into multiple reference regions of pixels, wherein the prior video frame is derived from the previously stored data; and determining a first reference region within the multiple reference regions of pixels that is most like a selected current region within the multiple current regions of pixels, the determination being utilized to determine the noise.

 The present invention can also be viewed as providing a method for providing
20 automatic reduction of noise for video encoding. In this regard, the method can be broadly summarized by the following steps: filtering noise from currently received video data; combining the filtered data, wherein the combining step is dependent upon a category of the noise; providing a weighted average of a current field derived from the combined

filtered data and a prior field, wherein the prior field is derived from previously combined and filtered data that has been previously stored, the weighted average being determined by pixel motion between the current field and the prior field; separating a current video frame into multiple current regions of pixels and separating a prior video frame into multiple reference regions of pixels, wherein the prior video frame is derived from the previously stored data; and determining a first reference region within the multiple reference regions of pixels that is most like a selected current region within the multiple current regions of pixels, the determination being utilized to determine the noise.

Other systems and methods of the present invention will be or become apparent to one with skill in the art upon examination of the following drawings and detailed description. It is intended that all such additional systems, methods, features, and advantages be included within this description, be within the scope of the present invention, and be protected by the accompanying claims.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention can be better understood with reference to the following drawings.

The components of the drawings are not necessarily to scale, emphasis instead being

placed upon clearly illustrating the principles of the present invention. Moreover, in the
5 drawings, like referenced numerals designate corresponding parts throughout the several views.

FIG. 1 is a typical computer or processor-based system in which the noise
reduction system of the present invention may be provided.

10 FIG. 2 is a block diagram further illustrating the video compression module of FIG.
1.

FIG. 3 further illustrates processing performed by the video input module of FIG. 2.

FIG. 4 is a block diagram further illustrating the temporal filter of FIG. 3.

FIG. 5 is a graph of a basic curve illustrating the relationship between pixel motion
15 and the weight of the previous frame upon the current frame.

FIG. 6 is a graph that provides an example of a family of curves parameterized
by χ from which one curve is to be selected based on estimated noise-level.

FIG. 7 is a graph that illustrates the window function for the value $L = 0$, wherein
the window function is utilized for motion-based limiting.

20

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

The noise reduction system of the present invention can be implemented in software, firmware, hardware, or a combination thereof. Specifically, a portion of the system may be implemented in software that is executed by a computer, for example, but not limited to, a server, a personal computer, workstation, minicomputer, or mainframe computer.

The software-based portion of the noise reduction system, which comprises an ordered listing of executable instructions for implementing logical functions, can be embodied in any computer-readable medium for use by, or in connection with, an instruction execution system, apparatus, or device such as a computer-based system processor containing system, or other system that can fetch the instructions from the instruction execution system, apparatus, or device and execute the instructions. In the context of this document, a “computer-readable medium” can be any means that can contain, store, communicate, propagate or transport the program for use by or in connection with the instruction execution system, apparatus or device.

The computer-readable medium can be, for example, but not limited to, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, device, or propagation medium. More specific examples (a non-exhaustive list) of the computer-readable medium would include the following: an electrical connection (electronic) having one or more wires, a portable computer diskette (magnetic), a random access memory (RAM) (magnetic), a read-only memory (ROM) (magnetic), an erasable programmable read-only memory (EPROM or Flash memory) (magnetic), an optical fiber

(optical), and a portable compact disk read-only memory (CD ROM) (optical). Note that the computer-readable medium could even be paper or another suitable medium upon which the program is printed, as the program can be electronically captured, via for instance, optical scanning of the paper or other medium, then compiled, interpreted or otherwise processed in a suitable manner, if necessary, and then stored in a computer memory.

Referring now to the drawings, wherein like reference numerals designate corresponding parts throughout the drawings, FIG. 1 is a typical computer or processor-based system 202 in which the noise reduction system of the present invention may be provided. The computer system of FIG. 1 generally comprises a processor 204 and a memory 206, having an operating system 208. Herein, the memory 206 may be any combination of volatile and nonvolatile memory elements, such as random access memory or read only memory. The processor 204 accepts instructions and data from the memory 206 over a local interface 212, such as a bus(es). The system also includes an input device(s) 214 and an output device(s) 216. Examples of input devices 214 may include, but are not limited to a serial port, a scanner, or a local access network connection. Examples of output devices 216 may include, but are not limited to, a video display, a Universal Serial Bus, or a printer port. Generally, this system may run any of a number of different platforms and operating systems, including, but not limited to, Windows NT™, Unix™, or Sun Solaris™ operating systems.

A video compression module 220 is located within the computer system 202 and is connected to the system via the local interface 212. Noise reduction is performed by the present noise reduction system via use of the video compression module 220 and logical

devices located therein. Further description of the video compression module 220 is provided hereinbelow with reference to FIG. 2. A communication interface 218 is also located within the computer system 202 for receiving video that is transmitted to the computer system 202.

5 Functionality for defining performance by the noise reduction system of the present invention, which is not defined and executed by the video compression module 220, is defined by software 205 that resides in memory 206. Of course, other functions for typical computer operation are defined within the memory 206 as well. Functionality performed by the present noise reduction system is defined hereinafter in further detail.

10 In addition, it should be noted that the noise reduction system need not be located within a computer system such as that illustrated by FIG. 1. Instead, the noise reduction system may be provided within an application specific integrated circuit (ASIC), wherein the ASIC is located on a circuit board. Further, the present system may be utilized in combination with the compression of any type of video.

15 FIG. 2 is a block diagram further illustrating the video compression module 220 of FIG. 1. As is shown by FIG. 2, the video compression module 220 comprises a video-input module (VIM) 222, a motion estimation unit 224, an encoder unit 226 and a video software processor 228. Each of these portions of the video compression module 220 is described in further detail herein below. It should be noted that the video software
20 processor 228 is preferably utilized to provide interaction between the VIM 222, the motion estimation unit 224, and the encoder unit 226, as is described in further detail hereinbelow.

Video data to be compressed by the video compression module 220 is fed to the VIM 222. As is known in the art, individual video fields are provided in succession to allow proper compression of video data. The video field is provided to the VIM 222 via a numerical representation that represents brightness (luma or Y) and color (chroma, or Cb and Cr) of pixels within the video field, otherwise known as YUV or YCbCr data. In accordance with the present embodiment, each luma or chroma value is represented as an 8-bit digital number, and may have a value between 0 and 255, inclusive. Of course, each luma or chroma value may be represented as a digital number having more or less bits and may have a different value, however, the aforementioned is assumed for demonstration purposes.

Processing performed by the VIM 222 is described by FIG. 3, which is a block diagram that further illustrates the VIM 222 of FIG. 2. As is shown by FIG. 3, video field data that is received by the VIM 222 is received via an input interface 252. Configuration of the video data may be performed in accordance with ITU-R Recommendation 656, which, as is known in the art, is an adopted standard used for uncompressed digital video. This standard uses a color sampling density that is half that of the brightness, known as “4:2:2” sampling.

From the input interface 252 the video field data is transmitted in parallel to a median filter 254 and a linear filter 256. The median filter 254 removes impulsive noise from the video field data. The linear filter 256 is preferably a low pass filter that attenuates high frequency noise and limits bandwidth. It should be noted that the order of filtering may differ in accordance with alternate embodiments of the invention. As an example, it

may be beneficial to perform filtering by the median filter 254 first and by the linear filter 256 second.

After filtering has been performed by the median filter 254 and the linear filter 256, the remaining filtered video field data is transmitted to a blender device 258. The blender device 258 combines the remaining filtered video field data. Blending of the filtered video data may be performed in different percentages wherein a higher, lower, or equal amount of median filter 254 filtered video field data is utilized, in comparison to linear filter 256 filtered video field data. The percentage of filtered video field data from one filter as opposed to the other filter may be dependent upon the type of noise present in the video field data. In addition, the blender device 258 may be used to utilize only data that has been filtered by the linear filter 256 alone or the median filter 254 alone. Furthermore, the blender device 258 may be used to achieve a specified linear filter response in areas where the median filter 254 acts to pass the input through to the output, and to achieve a noise spreading operation where the median filter 254 produces an output different from the input.

Once combining is completed by the blender device 258, the video field data is transmitted to a temporal filter 262. One input to the temporal filter is from a local memory 272. The local memory 272 has stored therein pixels of previously filtered and combined video field data, and is therefore also connected to the output of the temporal filter 262. The stored pixels are used for comparison and filtering purposes by the temporal filter 262, as is described in detail hereinbelow.

Pixels of the previously filtered and combined video data field are transmitted to the temporal filter 262 from the local memory 272. The temporal filter 262 compares pixel

values within the current filtered and combined video data field to spatially co-located pixel values from a previously filtered and combined video data field. The stored previously filtered and combined video data field is preferably of the same spatial parity (top field or bottom field) as the current filtered and combined video data field, although it is possible that pixels of a frame other than the immediately prior video data frame may be utilized.

During comparison performed by the temporal filter 262, the temporal filter 262 also performs averaging of pixel values. Averaging by the temporal filter 262 is performed by averaging the pixel values for a specific location in the current video data frame with the stored pixel values for the same location in the prior frame. Specifically, the luma (brightness), respectively chroma (color), values in the current video data field are averaged with the co-located luma, respectively chroma, values in the stored previously filtered and combined video data field. The values for luma and chroma are preferably represented by an eight-bit number. Of course, luma and chroma may be represented by a number that is represented by additional or fewer bits. As is known in the art, values for luma and chroma are represented by numerical values; therefore, averaging simply results in another numerical value.

FIG. 4 is a block diagram further illustrating the temporal filter 262 of FIG. 3. The temporal filter 262 provides a weighted average of a previously filtered field and a current field. A conversion device 264, buffer 266, horizontal lowpass filter 268 and temporal filter memory 269 are utilized to determine how much weight of a previous filtered field should be used in averaging. The weight applied to the previous filtered field is represented by the variable M, and the weight applied to the current field is represented by

1-M. As is shown by FIG. 4, the pixel values of the current field, after filtering and combining, are subtracted from the pixel values for the previous field, wherein the pixel values for the previous field are stored within the local memory 272 (FIG. 3). The absolute value of the derived difference in pixel values is then calculated by a conversion device 264. Successive absolute difference values are stored within a buffer 266, which receives the absolute value differences. Alternatively, the absolute difference values may be stored within the local memory 272 (FIG. 3).

A horizontal low pass filter 268 provides a weighted average of successive pixel absolute differences. Specifically, the horizontal low pass filter 268 provides finite impulse response filtering on the successive pixel absolute differences. The number of pixel absolute difference values may differ, however, in accordance with the preferred embodiment of the invention, five pixel absolute difference values, associated with a five-pixel region, are utilized for averaging purposes. The resulting value after finite impulse response filtering is denoted by the variable D, which is used as a measure of pixel motion.

Intuitively, if an object has either moved into or out of the five-pixel region associated with the value D during the time interval between the current field and the previous field, the value of D will be large. In such a case, the desired behavior of the temporal filter 262 is to select the current field, representing an unfiltered or only lightly filtered current version of the object represented by the five-pixel region. Selection of the current field is done by ensuring that M is low when D is high. Blending the previous and current fields with a high value of M in the presence of excessive motion may result in “ghosting,” in which two copies of the moving object would be visible in the current field, and/or blurring, in which the object appears to be smeared in the direction of motion.

Alternatively, if there is no motion during the time interval between the current field and the previous field, it is desirable to blend the current field with the previously filtered field so as to reduce the level of any noise that may be present through filtering. This kind of blending is accomplished by ensuring that M is high (preferably, near to one, or unity) when D is low. Experimentation shows that five pixel absolute difference for the weighted average represents a good trade-off between generating a high D value in the presence of moving objects and a low D value in the presence of only noise of low or moderate amplitude. It should, however, be noted that other sized regions may be utilized.

The variable D , or pixel motion, is used to determine the weight of the previous field in the weighted averaging process. In other words, the value of the variable D is utilized to determine what percentage of the previous field is incorporated into the present field. To make this determination, the value of the variable D is searched on the temporal filter memory 269. The temporal filter memory 269 comprises predefined values, represented herein by the variable M , which are located within a table. This table is sometimes referred to hereinbelow as a nonlinear transfer function (NLTF).

The temporal filter 262 utilizes the derived value of M to combine the current field data with the previous field data. Therefore, as mentioned hereinabove, a high value for M results in a higher percentage of the previous field data being combined with the present field, thereby resulting in combined video field data that is very similar to the video field data from a prior field. Clearly, a low value of M is desired when there is excessive motion represented by the current and prior video data field.

Conceptually, given the above description, $M \in [0,1]$, that is M may take any value between zero and one, inclusive. Preferably, in accordance with the present embodiment,

M may take one of 257 values in the set $\left\{\frac{0}{256}, \frac{1}{256}, \dots, \frac{256}{256}\right\}$. Furthermore, D may range

from 0 to 255, inclusive. The values of M are utilized by the present noise reduction system to transform a pixel motion level into a recursive filter coefficient utilized by the temporal filter 262 that provides a balance between noise reduction and motion artifacts.

- 5 A basic curve illustrating the relationship between pixel motion (D) and the weight applied to the previous field (M) is shown by FIG. 5. As is shown, the general form of the curve demonstrates that with high amounts of motion (high D), there is a low weight of the previously filtered field utilized in filtering the present field.

The following example is provided to further demonstrate the relationship between
10 the pixel motion (D) and the weight of the previous frame upon the current frame (M) in deriving resulting current video frame pixels. In accordance with the first example, a slowly moving video camera is utilized to record a stationary scene such as a room, while a large amount of noise affects the recording. This noise may be attributed to numerous different elements such as, but not limited to, low illumination levels, interference from
15 other electronic devices, or the power source to the video camera. Due to the amount of noise affecting recording of the room, and the minimal amount of movement during recording, it is desirable to perform a large amount of filtering of noise to remove the abovementioned interference, *i.e.* to utilize a high percentage of video field data from the prior video field pixels (high M) when calculating the resulting current video field pixels.

20 By utilizing a high percentage of the prior video field pixels (high M), the noise encountered by the present video field data during recording is averaged out with the previously filtered and stored pixel values of the previous video field. Further, since the camera is moving slowly, there is a large amount of similarity between the current frame

and the prior frame so that a high value for M does not introduce excessive blurring or “ghost” effects.

By contrast, in accordance with a second example, the video camera is recording fast paced action such as a car race while minimal noise is encountered. Due to the vast amount of motion recorded, there is a large difference between the present video field and the prior video field. Specifically, due to the vast amount of motion, the co-located pixel values for the prior video field are quite different from the pixel values for the present video field. Therefore, it is not desirable to utilize a high percentage of the prior video field pixels during calculation of the resulting current video field pixels. Further, since there is minimal noise, utilization of a high percentage of the prior video field pixels for averaging is not necessary.

Unfortunately, the temporal filter 262 has certain limitations in its ability to distinguish moving objects from noise. This results in a danger of blurring or ghosting of moving objects unless a very conservative approach to setting the values in temporal filter memory 269 is used. Use of a conservative approach has the effect of not providing as much noise reduction as can otherwise be achieved. To address this limitation, prior art temporal filters utilize either improved pixel motion detection to help select the proper D value, or even motion compensated filtering to allow filtering in the presence of motion. However, each of those methods requires additional hardware resources. Furthermore, noise levels can change over time as the video source changes, and a temporal filter memory 269, in which the M vs. D curve is static, cannot provide the best noise reduction at all noise levels.

As a result, the present noise reduction system partially circumvents these limitations by dynamically controlling the M entries in the temporal filter memory 269 table (NLTF) over time in response to incoming video fields. In accordance with the preferred embodiment of the invention, the dynamic control is performed once per field of video, although it can be done more or less frequently. The noise reduction system performs this dynamic control by using noise and motion level information generated by the motion estimation unit 224 and statistics generator 267 during the course of operation, as is described herein below.

The principles behind dynamic control are as follows. The NLTF may be set at an optimum level for reducing a given level of noise, if an estimate of the prevailing noise level can be made during the course of operation of the noise reduction system. Using the noise estimate, a basic NLTF is selected, without considering whether motion is present, which would interact with the temporal filter 262, creating the above mentioned ghosting and blurring artifacts. This generally results in a temporal filter 262 that may cause motion artifacts, but that reduces noise very well. To address the issue of motion artifacts, a second step is performed for estimating the motion level and conditionally reducing the amount of filtering by modifying the basic NLTF. An example of a method for determining noise and motion estimates and using the estimates to dynamically control the temporal filter 262 is described below. It should be noted that other methods for determining noise and motion estimates are also possible.

The following describes a method for providing noise level estimation and also describes how the method is used to select a basic NLTF. In addition, the method by

which the motion level is estimated is provided and how the motion level estimate is applied for modification of the NLTF.

Estimating Noise Level and Computing a Basic NLTF

5 Returning to FIG. 2, one purpose of the motion estimation unit 224 is to provide information regarding moving objects to the encoder unit 226 for video compression purposes. In addition, information that is also stored within the local memory 272 as a result of the motion estimation process is used by the present noise reduction system to estimate noise levels. These estimated noise levels are then further utilized by the noise
10 reduction system to determine values of M that are stored within the temporal filter memory 269, thereby providing the basic NLTF. This process is described in detail herein below.

 The motion estimation unit 224 first separates a first video frame, labeled as the current frame, into regions of pixels having a predefined size and number of pixels.
15 Preferably, a sixteen pixel by sixteen pixel region is selected, although a region of different size may instead be selected. The motion estimation unit 224 then examines a second video frame, labeled as the reference frame, and compares the region for the current frame to determine what same-sized area of the reference frame is most similar to the current video frame region. Examination of the reference frame is performed by searching for a
20 same-sized area having similar pixel values within the reference video frame. Since these pixel values are stored within the local memory 272, examination of the reference video frame is performed by searching the local memory 272. Numerous means may be utilized

to perform this search, including, but not limited to, hierarchical motion estimation, phase-correlation, and full search motion estimation.

Two pieces of information are produced by the examination process for each region in the current frame. The first piece of information, known as a motion vector, is simply a vector of integers representing a spatial displacement in the horizontal and vertical direction of the most similar region in the reference frame, with respect to the region in the current frame. The motion represented by this integer vector may be of higher precision than integer pixel displacements. Specifically, with reference to the preferred embodiment of the invention, the motion represented is of half-pixel precision.

The second piece of information produced by the examination process is a measure of the goodness of fit between the region in the current frame and the most similar region in the reference frame. This measure is denoted herein as the SAD, which is the sum of absolute pixel differences between the region in the current frame and the most similar region in the reference frame. Other measures, such as mean squared error, may be incorporated into the noise level estimation procedure.

The difference in pixel luma values between the 16x16 region in the current video frame and the similar 16x16 region in the reference frame may be attributed to numerous elements, one of which is noise. Therefore, since the motion estimation device 224 determines similar pixel regions, as has been described hereinabove with reference to the determination of motion, and since the difference between these regions may be attributed to noise, the motion estimation device 224 may be utilized for estimating noise.

Specifically, to estimate noise, the present system keeps track of the sum of absolute differences of pixel values for each current region / similar reference region pair.

The sum of absolute differences provided by the motion estimation device 224 is also referred to herein as a mean absolute difference (MAD) score. Therefore, in accordance with the present illustration of use of a 16x16 pixel region, each region is associated with one MAD score.

5 For each current / reference frame pair, the lowest MAD score is determined from all computed MAD scores. In an alternative embodiment, the lowest MAD score that is not zero is determined from all computed MAD scores. The lowest MAD score is referred to hereinafter as a *minMAD* score. Since incurred noise is different for each video frame, the *minMAD* score is first temporally smoothed. In accordance with the preferred
10 embodiment of the invention, this first smoothing is done with an infinite impulse response filter. A second nonlinear smoothing is performed over a series of successive frames, preferably, although not limited to, three frames. In the second smoothing the lowest *minMAD* score over the three frames is determined. Since the *minMAD* is indicative of the best match between a region from a current frame and a same-sized area in a reference
15 frame and since there is likely to be at least one good matching region between two frames in a video scene, the *minMAD* is a good indicator of noise.

A noise-based determination of the basic NLTF using *minMAD* is performed by using the following parameterized function:

$$20 \quad TF(D) = \min \left\{ \alpha, \beta \times e^{(-0.5((D+8)/(X+3))^2)} \right\} \quad (\text{Eq. 1})$$

The variable D is the output of the horizontal lowpass filter 268 in FIG. 4. Hereinabove, the variable α limits the infinite impulse response coefficient, the variable β determines

the area of the M vs. D curve, and the variable χ determines the rate at which the M vs. D decays to 0 as D increases. In accordance with the preferred embodiment of the invention, the values of α and β are fixed at 240 and 325, respectively, and the value of χ may be one of several possible values, referred to herein as temporal filter 262 levels. Specifically, there are 31 temporal filter levels (0,1,...,30) in accordance with the preferred embodiment. One skilled in the art will appreciate that other values and levels may be utilized. An example of a family of curves parameterized by χ is shown by the graph of FIG. 6.

Functional form of the curve parameterized by χ may be determined by computer-based fitting to a hand-tuned NLTF. Other functional forms of similar effectiveness may also be utilized. Similarly, many different families of curves may be derived from a suitably parameterized functional form without deviating from the intent of the present invention.

The temporally and nonlinearly smoothed *minMAD* noise estimate is converted to the temporal filter level χ , using the following linear mapping:

$$\chi = \frac{3}{4} \times noise + 4 \quad (\text{Eq. 2})$$

Estimating Motion Level and Modifying the Basic NLTF to Form a Final NLTF

Unfortunately, if there is a large amount of video motion or an extreme difference between fields, such as during a change of scenes, it is not desirable to utilize a prior video field in the derivation and filtering of a current field. If the prior video field were used in spite of these extreme differences between the present field and the prior field, remnants of the prior field would be incorporated into a completely different video scene, also referred herein above as ghosting artifacts. Therefore, the VIM 222 is utilized to determine the amount of scene motion, denoted herein as the motion level, in order to mitigate ghosting artifacts by limiting the amount of temporal filtering performed during excessive motion and scene changes. This motion level is used to modify the basic NLTF to prevent these artifacts. The following provides an example of a method utilized to determine motion level and illustrates use of the method to modify the basic NLTF.

For the purposes of noise reduction, motion level may be represented by the following equation:

$$motion\ level = \frac{SAD_{hist}}{gSAD} \quad (Eq. 3)$$

In this equation, $gSAD$ is the sum of absolute pixel differences for the entire field and is defined below, while SAD_{hist} represents the sum of absolute pixel differences not for the entire field, but only for those pixels whose absolute pixel difference exceeds a *threshold*. Labeling pixel i in the current field $y_n(i)$ and pixel i in the previous spatially co-located field $y_{n-1}(i)$, SAD_{hist} is defined as:

$$SADhist \equiv \sum_{i=0}^{N-1} |y_n(i) - y_{n-1}(i)| \times I(|y_n(i) - y_{n-1}(i)| > threshold), \quad (\text{Eq. 4})$$

where $I(\text{condition})$ is the indicator function, and equals 1 if condition is true and 0

5 otherwise. However, since *SADhist* is not computed herein, it is estimated. *SADhist* is estimated using measurements computed by the statistics generator 267, specifically *gSAD*, *maxZoneSAD* and *histCount*, which are defined and described hereinbelow.

For each incoming video field, the statistics generator 267 computes a first means of measurement to be the sum of absolute pixel differences between the current incoming
10 field and the previous spatially co-located filtered field from local memory 272. This sum is labeled *gSAD* hereinbelow. Labeling pixel i in the current field $y_n(i)$ and pixel i in the previous spatially co-located field $y_{n-1}(i)$, *gSAD* is given by the equation:

$$gSAD = \sum_{i=0}^{N-1} |y_n(i) - y_{n-1}(i)|, \quad (\text{Eq. 5})$$

where N is the number of pixels in a field.

15 A second means of measurement produced by the statistics generator 267 is a *maxZoneSAD*, which is the maximum sum of absolute differences between the present field and the previous spatially co-located field, over a zone. A zone is a rectangular region of the field that is programmable in size by the present noise reduction system. As a non-limiting example, the zone may be sixteen pixels by six pixels in size, although it
20 may be of different proportions. It should be noted that the larger the size of the zone, the more difficult it is to detect small moving objects within a video. After designation of the proportions of a zone, the sum of absolute differences over each individual zone is

determined. The value of the sum in the zone with the maximum sum is then designated as $maxZoneSAD$. Assuming there are M zones labeled $Z_m, m = 0, \dots, M-1$, $maxZoneSAD$ is given by the equation:

$$maxZoneSAD = \max_{m=0, \dots, M-1} \left\{ \sum_{i \in Z_m} |y_n(i) - y_{n-1}(i)| \right\}. \quad (\text{Eq. 6})$$

5

A third means of measurement produced by the statistics generator 267 is a $histCount$, being the number of times the pixel absolute difference is above a specified threshold. The threshold is programmable, as is described in detail hereinbelow.

The following equation estimates $SADhist$ using $gSAD$, $histCount$ and $maxZoneSAD$:

$$SADhist \approx (A \cdot gSAD/N_{\text{field}} + B \cdot maxZoneSad/N_{\text{zone}} + C) \cdot histCount \quad (\text{Eq. 7})$$

where N_{field} is the number of pixels in a field, N_{zone} is the number of pixels in a zone, and A , B , and C are regression coefficients. Since luma and chroma statistics exhibit different levels of sensitivity to motion, different histogram threshold and different sets of regression coefficients A , B and C for luma and chroma statistics are used. The following provides an example of values that may be utilized in the above-mentioned equation 7 for purposes of illustration. It should be noted, however, that other values may be utilized.

A histogram threshold of 20 for luma and 10 for chroma works well for low to medium noise. Regarding the regression coefficients, for luma, $A = 0.557838595$, $B = 0.126760734$ and $C = 21.35959058$; and for chroma, $A = 0.196923005$,

B = 0.092330113 and C = 12.00067648. It should be reiterated that *SADhist* is not computed in the actual noise reduction system.

Combining equations 3 and 7 results in the following equation that is utilized to derive a measure of motion level between fields:

5

$$motion\ level = \frac{(A \cdot gSAD/N_{field} + B \cdot \max ZoneSad/N_{zone} + C) \cdot histCount}{gSAD} \quad (Eq. 8)$$

It should be noted that a high resulting value for motion level indicates that there is a large amount of motion between the present field and the previous field, and that consequently, less temporal filtering should be applied.

10

Given the above estimate of motion level, motion-based limiting of the M values is performed, in accordance with the preferred embodiment of the invention, by using the following windowing function:

15

$$W(D, L) = \frac{\exp(a - b \exp(D - L))}{\exp(a - b)} \quad (Eq. 9)$$

Herein, the variable *D* is pixel motion, as defined hereinabove, and the variable *L* is a temporal filter 262 limit value. In accordance with the present example, the constants *a* and *b* are selected as *a* = 0.001512 and *b* = 2.17x10⁻⁸. The temporal filter limit *M* is then related to the motion level estimate using the following equation:

20

$$L = (24/motion\ level) - 18 \quad (Eq. 10)$$

FIG. 7 is a graph that illustrates the window function for the value $L = 0$. Changing the value of L moves the transition to the left or right. Respective of the fact that this window function will be multiplied by the basic NLTF to form a final NLTF, it is evident that it serves to leave the basic NLTF unchanged below a first threshold, to reduce the basic NLTF to 0 above a second threshold only slightly larger than the first threshold, and to provide a smooth transition from unchanged to 0 between the first and second thresholds. One skilled in the art will appreciate the abovementioned may be achieved via other means, including a different window function or other means of modifying the basic NLTF.

Final NLTF Based on Noise and Motion Level Estimates

The following equation is utilized to compute M values given a temporal filter level χ and a motion limit value L :

$$M = \min\{A, W(D, L) \times \text{Bexp}(-0.5((D+8)/(\chi+3))^2)\} \quad (\text{Eq. 11})$$

In addition to the above considerations regarding the automatic changing of the temporal filter memory 269, it is beneficial to completely disable temporal filtering (*i.e.* set all M values to 0) immediately after a change of scene is detected and then to resume normal operation one frame later.

In the presence of high noise levels, the temporal filter 262 alone may be insufficient for noise removal. An example of such high noise levels includes “snow” that

appears when a television is tuned to an unused channel. In such a situation, the linear filter 256 may be programmed to perform lowpass filtering to further reduce spatial bandwidth of the incoming video field, thus reducing high-frequency noise.

Unfortunately, lowpass filtering also reduces picture resolution, thereby resulting in this option being used in extremely noisy circumstances in which the tradeoff between resolution and decreased noise is acceptable.

A family of sets of coefficients for the linear filter 256, for each of the horizontal luma, vertical luma, horizontal chroma, and vertical chroma filters, is preferably utilized. As an example, there may be a family of 16 sets of coefficients. Given a linear filter 256 level S in the range of 0 to 15, inclusive, the horizontal luma, vertical luma, horizontal chroma, and vertical chroma filters have a cutoff of approximately $0.5 + 0.05(16 - S)$ in normalized frequency. Preferably, the luma and chroma filters are both 9-tap horizontal FIR linear phase filters, with S controlling all filters. Since one skilled in the art would understand how to design and/or develop such filters, further discussion is not provided herein.

A change in the linear filter 256 level S is readily perceived in the output video as a change in the “softness” or level of detail. Unfortunately, if the value of S changes quickly from field to field, the results of the change may be quite displeasing. To avoid this change S is smoothed. Preferably, a first-order infinite impulse response (IIR) filter is utilized for smoothing purposes, although other filters may be substituted. In addition, it is possible and beneficial to reduce the frequency with which a change in the linear filter 256 level is updated from every field to a lower frequency such as every third frame.

The noise level estimate described hereinabove is converted to the linear filter level S using linear mapping. As an example the following linear mapping may be utilized, which also shows IIR smoothing by a factor of α .

$$S = \alpha S_{\text{old}} + (1 - \alpha)(0.324166 \cdot \text{noise} - 1.792115) \quad (\text{Eq. 12})$$

5 It should be noted that, unlike the temporal filter 262, the linear filter 256 does not need to be turned off at a scene change.

It should also be noted that dynamic control of the linear filter 256 and the temporal filter 262 can be performed separately or simultaneously, and that the dynamic control of the temporal filter 262 can be performed on the basis of both noise level and motion level
10 or on the basis of one or the other. To provide maximum benefit, the preferred embodiment of the invention uses dynamic control of both the linear filter 256 and the temporal filter 262 on the basis of both noise level and motion level estimates.

It should be emphasized that the above-described embodiments of the present invention, particularly, any “preferred” embodiments, are merely possible examples of
15 implementations, merely set forth for a clear understanding of the principles of the invention. Many variations and modifications may be made to the above-described embodiment(s) of the invention without departing substantially from the spirit and principles of the invention. All such modifications and variations are intended to be included herein within the scope of this disclosure and the present invention and protected
20 by the following claims.